ESTIMATION AND PREDICTION WITH THE MIXED LOGIT MODEL: ANALYSIS OF SOME CONFOUNDING EFFECTS

Elisabetta Cherchi
CRiMM - Dipartimento di Ingegneria del Territorio
Facoltà di Ingegneria - Università di Cagliari
Piazza d'Armi, 16 - 09123 Cagliari, Italia
Tel: 39 70 675 5274; Fax: 39 70 675 5261;
E-mail: echerchi@unica.it

and

Juan de Dios Ortúzar
Departamento de Ingeniería de Transporte
Pontificia Universidad Católica de Chile
Casilla 306, Cod. 105, Santiago 22, Chile
Tel: 56-2-354 4822; Fax: 56-2-553 0281;
E-mail: jos@ing.puc.cl

ABSTRACT

The Mixed Logit (ML) model utility function is characterised by an error term with two components; one, designed to capture white noise effects, allows obtaining a basic logit probability and the other, designed to achieve a fairly general covariance matrix, has a distribution that can be freely chosen by the modeller. This flexible structure is at the roots of the great popularity of the ML but is also the cause of a major problem, as several confounding effects among correlation, preference, response unobserved heterogeneity and heteroskedasticity, might occur.

We analyse the problem raised by the confounding effects in the ML and their relation with the trade-off implicit in the typical linear in the parameters with additive disturbances structure. In particular, we refer to three assumptions common to almost any discrete choice model: (a) that individuals evaluate alternatives using a compensatory rule; (b) that unobserved attributes might be accounted for by an additive random structure, and (c) that only the difference between alternatives matters. Using simulated data we demonstrate that confounding effects are intrinsic to the "true" role of any random term: distinguishing or adding commonality to pairs of alternatives. This is not as obvious as it may appear, and thinking of error terms as "proxy" for unobserved omitted structures helps to understand their "true" role and aids in the search for better specifications.

1. INTRODUCTION

The Mixed Logit (ML) label is applicable to any model whose choice probabilities can be expressed as the integral of standard logit probabilities over a density of parameters. The ML utility function is characterised by an error term with two components. The first allows to obtain the logit probability (and thus has the usual *iid* extreme value type I distribution). The second has a distribution which can be freely chosen by the modeller, depending on the phenomenon s/he needs to reproduce. The most general way of thinking the ML model is to specify the second error term as the product of a vector of attributes and its relative taste parameters. Depending on whether the attributes included in the second part of the error term are known, we can obtain a random parameters specification or an error components model.

As most discrete choice models, both ML forms assume that individuals evaluate each alternative using a compensatory rule, i.e. they actually trade-off among attributes. It has been shown that there are situations where compensatory rules do not hold (Cantillo and Ortúzar, 2005), however most applications in transport assume that individual behaviour follows compensatory rules, as this provides a systematic utility form that is very easy to handle. As the modeller does not know the "true" utility (attributes and form) s/he tries to explain its unknown part by means of random elements. The most widely used specifications is the LPAD structure (i.e. linear in the parameter with added disturbances) that, from a formal point of view, extends the compensatory rule also to the random terms, the role of which is actually compensating for the omission of relevant attributes. Interestingly, we found that this structure (i.e. the trade-off between systematic and random parameters) is at the basis of several confounding effects that can appear in the estimated models.

Several studies have raised an alarm regarding the difficulty of distinguish different substitution patterns in a ML, but have mainly referred to the scale problem. For example, Munizaga and Alvarez (2001) show that it is not straightforward to reproduce a Nested Logit (NL) model with the ML because of the inherent heteroskedasticity of the latter. Hensher and Greene (2003) suggest that heterogeneity is a special type of correlation, amongst choice situations. Bhat and Castelar (2002) point out to the likely interactions among inter-alternative error structures, RP/SP scale differences, unobserved heterogeneity and state dependence. Hensher (1998) found that specific alternatives may produce spurious correlation if unobserved heterogeneity is ignored, and Swait and Bernardino (2000) show that taste heterogeneity may be potentially confounded with heteroskedasticity and preference trees.

In this paper we analyse the confounding effect between correlation among alternatives, heterogeneity in tastes and heterogeneity in response, and highlight their relation with the trade-off implicit in the LPAD structure. Moreover, as discrete choice models work on differences between alternatives, serious confounding effects might occur also between different alternatives. The interest for these effects is not only theoretical. Strictly related to it there is a problem of understanding the postulated error structure, interpreting the results obtained and correctly using the model in forecasting. Using simulated data, we provide empirical evidence for our theoretical discussion.

The rest of the paper is organised as follows. In section 2 we review the basics of the ML and set out the formal relation between the typical structure of discrete choice models and some confounding effects. Section 3 describes the dataset used for our empirical tests. Section 4 reports the results obtained by estimating several models under different assumptions about the underlying dataset, and discuss several confounding effects that might arise and their effects in prediction. Finally section 5 summarises our conclusions.

2. CONFOUNDING EFFECTS IN THE MIXED LOGIT MODEL

The ML is any model whose choice probabilities can be expressed as the integral of standard logit probabilities, evaluated at parameters α, over a density of parameters:

$$P_{qj} = \int \frac{e^{\alpha_{qj} L_{qj}}}{\sum_{i \in I_q} e^{\alpha_{qi} L_{qi}}} f(\alpha_{qj} | \Omega) d\alpha \tag{1}$$

where P_{qj} is the probability of individual q choosing alternative j among any i-th alternative belonging to her set (I_q) of available alternatives; Ω are the population parameters of the distribution and L_{qj} is any linear transformation of the relevant attributes for the individual q and alternative j. The only requirement for a ML model is that the random component of the individual utility has an additive GEV type 1 error, which generates the standard logit probability, while the vector of parameters α_{qj} can assume any desired distributions (or can be fixed). In fact, as demonstrated by McFadden and Train (2000) any well-behaved random utility (RUM) model can be approximated to any degree of accuracy by a ML model.

Let be, $L_{qj} = (k_j, \mathbf{x}_{qj}, \mathbf{z}_{qj})$ and $\alpha_{qj} = (\mathbf{b}_j, \boldsymbol{\mu}_{qj})$, and let's write, without loss of generality, the individual utility as:

$$U_{qj} = \underbrace{k_j + \mathbf{b}_j^{'} \mathbf{x}_{qj}}_{V_{qj}} + \underbrace{\boldsymbol{\mu}_{qj}^{'} \mathbf{z}_{qj}}_{\delta_{qj}} + \varepsilon_{qj} \tag{2}$$

where $\mathbf{x}_{qj}$ are vectors of known (by the modeller) characteristics of the alternative j for individual q; $\mathbf{b}_j$ is a vector of taste parameters fixed over the population; k_j is the specific constant for alternative j; $\mathbf{z}_{qj}$ is a vector of attributes that could be known (i.e. equal to $\mathbf{x}_{qj}$) or unknown, and $\boldsymbol{\mu}_{qj}$ a vector of unobserved taste parameters randomly distributed over the population for each alternative. Finally ε_{qj} is the additive GEV type 1 error; following Hensher and Green (2003), these can be interpreted as task-specific shocks to q's tastes.

It is well-known that the ML can assume two forms depending on whether the analyst knows the vector $\mathbf{z}_{qj}$ of attributes or not. In the first case (i.e. $\mathbf{z}_{qj} = \mathbf{x}_{qj}$), the typical random parameters (RP) model is obtained:

$$U_{qj} = k_j + \boldsymbol{\beta}_{qj}^{'} \mathbf{x}_{qj} + \varepsilon_{qj} \tag{3}$$

where $\beta_{qj} \approx f(b_j, \Omega)$ or, equivalently, $\mu_{qj} \approx f(0, \Omega)$ where f is any desired distribution. In the second case, instead, $\mathbf{z}_{qj}$ is unknown and assuming without loss of generality that $z_{qj}=1$ for all alternatives (or for groups of them), the error components (EC) model is obtained:

$$U_{qj} = \hat{k}_j + \mathbf{b}_j^{'} \mathbf{x}_{qj} + \boldsymbol{\eta}_{qj} \tag{4}$$

where $\boldsymbol{\eta}_{qj} = (\mu_{qj}, \varepsilon_{qj})$ and $\mu_{qj} \approx f(0, \Omega)$, with f being any desired distribution.

Whatever is the expression and the interpretation, both ML forms assume that individuals evaluate each alternative using compensatory rules, such that the effects of bad attributes can be compensated by the effect of good attributes. Although it has been shown that there are situations where compensatory rules do not hold (Cantillo and Ortúzar, 2005), most applications in the transport field still assume that individual behaviour follows compensatory rules, as this provides a very easy to handle form for the systematic utility.

Now, it is interesting to note that since an additive structure for the error terms is also assumed, although the compensatory behaviour only concerns the attributes in practice it also extends to the random terms. From a modelling point of view, this is perfectly acceptable; in fact, it is well known that the random terms account for all the effects that modellers are not able to observe and include as components the vectors of characteristics (Manski, 1977; McFadden, 1981); therefore, random terms are nothing more than attributes simply expressed differently by the modellers. This effect is evident in the case of the iid Gumbel white noise error and the EC structure, but it also applies to the RP structure.

Moreover, discrete choice models work on the basis of differences between alternatives; so it does not matter whether one attribute is included in alternative j, or in all the others except j, as long as the relative difference between alternatives does not change (Ben-Akiva and Lerman, 1985). This property of discrete choice models in conjunction with the compensatory rule is at the root of several confounding effects that can appear in the estimated models. It is worth highlighting that the confounding effects are implicit in the structure of the covariance matrix of the ML model. Nevertheless, understanding which effects can be confounded and what parameters are affected (and therefore wrongly estimated) is important to interpreting correctly the results obtained and correctly using the model in forecasting.

Firstly, it should be noted that an EC model is always able to reproduce an alternative specific RP structure, while the opposite might happen but it is not always true because the RP structure is constrained by the attributes associated to it. Secondly, since the EC model does not depend on attributes associated to alternatives, the following two structures are indeed equivalent:

$$\begin{array}{llllll} U_{q1} & = k_1 & + \mathbf{b}^{'}\mathbf{x}_{q1} & + \mu_q & + \varepsilon_{q1} & \\ U_{q2} & = k_2 & + \mathbf{b}^{'}\mathbf{x}_{q2} & & + \varepsilon_{q2} & \text{(5a)} \\ U_{q3} & = & + \mathbf{b}^{'}\mathbf{x}_{q3} & & + \varepsilon_{q3} & \end{array} \qquad \begin{array}{llllll} U_{q1} & = k_1 & + \mathbf{b}^{'}\mathbf{x}_{q1} & & + \varepsilon_{q1} & \\ U_{q2} & = k_2 & + \mathbf{b}^{'}\mathbf{x}_{q2} & + \mu_q & + \varepsilon_{q2} & \text{(5b)} \\ U_{q3} & = & + \mathbf{b}^{'}\mathbf{x}_{q3} & + \mu_q & + \varepsilon_{q3} & \end{array}$$

Although formally different the correlation between alternatives and EC structures are complementary. In both cases, in fact, we estimate the differences:

$$\begin{aligned} U_{q1}-U_{q2} &= k_1-k_2 &+\mathbf{b}'(\mathbf{x}_{q1}-\mathbf{x}_{q2}) &+\mu_q &+\varepsilon_{q1}-\varepsilon_{q2} \\ U_{q1}-U_{q3} &= k_1 &+\mathbf{b}'(\mathbf{x}_{q1}-\mathbf{x}_{q3}) &+\mu_q &+\varepsilon_{q1}-\varepsilon_{q3} \\ U_{q2}-U_{q3} &= k_2 &+\mathbf{b}'(\mathbf{x}_{q2}-\mathbf{x}_{q3}) & &+\varepsilon_{q2}-\varepsilon_{q3} \end{aligned} \quad (6)$$

Moreover, the following two structures can also give similar results:

$$\begin{aligned} U_{q1} &= \hat{k}_1 &+\mathbf{b}'\mathbf{x}_{q1} &+\hat{\mu}_q x_{i,q1} &+\varepsilon_{q1} \\ U_{q2} &= \hat{k}_2 &+\mathbf{b}'\mathbf{x}_{q2} & &+\varepsilon_{q2} \\ U_{q3} &= &+\mathbf{b}'\mathbf{x}_{q3} & &+\varepsilon_{q3} \end{aligned} \quad (7a) \qquad \begin{aligned} U_{q1} &= k_1 &+\mathbf{b}'\mathbf{x}_{q1} & &+\varepsilon_{q1} \\ U_{q2} &= k_2 &+\mathbf{b}'\mathbf{x}_{q2} &+\mu_q &+\varepsilon_{q2} \\ U_{q3} &= &+\mathbf{b}'\mathbf{x}_{q3} &+\mu_q &+\varepsilon_{q3} \end{aligned} \quad (7b)$$

where, the i-th attribute ($x_{i,q1}$) of the first alternative has a random parameter. In fact, in both case the "true" effect of the random terms is to differentiate alternative 1 from alternatives 2 and 3. In cases (7a) and (7b) major differences should be expected in the absolute value of the alternative specific constants (ASC), but not obviously in their relative difference. This is because accounting for random heterogeneity affects the mean of the difference between pairs of alternatives, but in the RP structure (equation 7a) $E(\hat{\mu}_q)=b_i$ (i.e. the mean value of the random parameter is the estimated fixed parameter of the associated attribute). Conversely, in the EC model (equation 7b) the above difference is accounted by the ASC. Therefore, we should have $\hat{k}_1=E(\varepsilon_{q1}-\varepsilon_{q3})$ and $\hat{k}_2=E(\varepsilon_{q2}-\varepsilon_{q3})$ in case (7a); and $k_1=E(\varepsilon_{q1}+\mu_q-\varepsilon_{q3})$ and $k_2=E(\varepsilon_{q2}-\varepsilon_{q3})$ in case (7b).

Finally, it is important to note that, in real cases the utility that individuals associate to each alternative is often complex and involves more attributes with different interactions (Galilea and Ortúzar, 2005), than those modellers are able to specify. Therefore, from one side it is likely to find some random structure which is significant; on the other side, it is difficult to be certain that the random structure that results significant from the estimation really represents only that specific effect. For example, let us suppose that individuals evaluate the alternatives according to this very simple expression: $U_{qj}=bx_{qj}+\gamma h_{qj}x_{qj}$, where b and γ are fixed parameters, but the modeller knows only the vector of attributes ($\mathbf{x}_{qj}$) and completely ignores h_{qj}. Because of the presence of the unknown attribute h_{qj}, which is distributed over the population with certain parameters (γ,Ω), it should not be surprising if random heterogeneity for x_1 is found although it is not really the true effect.

3. DATABANK

To test for the presence of the confounding effects discussed in the previous section a collection of datasets were simulated in which pseudo-observed individuals behaved in according to a choice process determined by the analyst following Ortúzar and Williams (1982). We used information (types of variables, mean and standard deviation) from a real sample collected in

1998 for mode choice among car, bus and train (Cherchi and Ortúzar, 2002). Three samples were generated each representing a different behaviour.

The first generated dataset (Dataset N.1) had individuals showing random taste heterogeneity in the travel time by car; all the other attributes were evaluated equally by all individuals in the sample and no correlation was present among alternatives. Thus, for each individual q, modal choices were generated according to the following utility function:

$$\begin{cases} U_{qj} = \beta_j + \beta_2 Ttime_{qj} + \delta_j \mu_{qj} Ttime_{qj} + \beta_3 C_{qj} + (1-\delta_j)\beta_4 Freq_{qj} + \beta_5 Wtime_{qj} + \varepsilon_{qj} \\ \delta_j = 1 \quad \text{if } j = \text{car; 0 otherwise} \\ \beta_j = 0 \quad \text{for } j = \text{bus} \\ \mu_{qj} \approx N(0, 0.35), \quad \text{with } \mu_{qj} \subset [-3, -0.001] \\ \varepsilon_{qj} \approx iid\ GEV\ I \end{cases} \tag{8}$$

where C_j is travel cost, $Ttime_j$ is in-vehicle travel time, $Wtime_j$ is walking time and $Freq_j$ the number of public transport (bus or train) vehicles per hour; μ_{qj} is the deviation of each individual travel time parameter from the mean value (β_2) and it has a censored distribution. The parameters had the following point values: β_3 = -0.6; β_2 = -0.8; β_4 = -0.22; β_5 = 0.6; $\beta_{j=car}$ = 2.0 and $\beta_{j=train}$ = -0.5. These values were taken from a MNL estimated with real data (i.e. the same sample from where the level-of-service (LOS) values had been taken), except β_2, the value of which was the mean of the random variable generated.

In the second dataset (Dataset N.2) the assumption was made that attributes were fixed over the sample but the public transport alternatives were perceived as correlated, according to the following utility function:

$$\begin{cases} U_{qj} = \beta_j + \beta_2 Ttime_{qj} + \beta_3 C_{qj} + (1-\delta_j)\beta_4 Freq_{qj} + \beta_5 Wtime_{qj} + \varepsilon_{qj} + (1-\delta_j)\eta_q \\ \delta_j = 1 \quad \text{if } j = \text{car; 0 otherwise} \\ \beta_j = 0 \quad \text{for } j = \text{bus} \\ \eta_q \approx N(0,1) \\ \varepsilon_{qj} \approx iid\ GEV\ I \end{cases} \tag{9}$$

where all the terms have the same meaning described above, and η_q is the error component introduced to generate correlation between bus and train.

The third sample was generated assuming a step-wise utility function for the travel time marginal utility, as illustrated in Figure 1. In particular, we assumed that whatever the length of the journey, the first ten minutes of travel would produce a marginal disutility equal to -0.6, while any minute above the 10^{th} was evaluated at the margin as equal to -0.1.

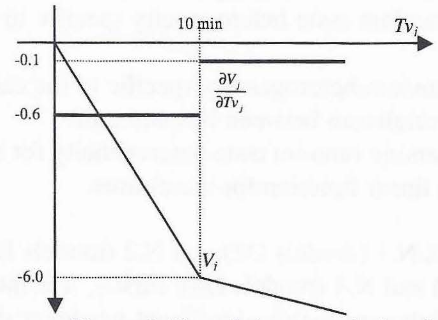

Figure 1: Travel time marginal and total utility

Dataset N.3 was then generated according to the following utility function:

$$U_{qj} = \beta_j + \beta_{21}Tt(\leq 10)_{qj} + \beta_{22}Tt(>10)_{qj} + \beta_3 C_{qj} + (1-\delta_j)\beta_4 Freq_{qj} + \beta_5 Wtime_{qj} + \varepsilon_{qj} \quad (10)$$

where all the variables have already been defined, except travel time *Tt*, which is given by:

$$Tt(\leq 10)_{qj} \begin{cases} = Ttime_{qj} & \text{for } Ttime_{qj} \leq 10 \text{ minutes} \\ = 10 & \text{for } Ttime_{qj} > 10 \text{ minutes} \end{cases}$$

$$Tt(>10)_{qj} \begin{cases} = 0 & \text{for } Ttime_{qj} \leq 10 \text{ minutes} \\ = (Ttime_{qj} - 10) & \text{for } Ttime_{qj} > 10 \text{ minutes} \end{cases}$$

Finally, the fourth dataset (Dataset N.4) was generated using a simple linear in the parameters and the attributes utility function, with no randomness in the parameters or correlation among alternatives.

The LOS variables used to build the simulated utility function were generated according to a truncated Normal distribution to guarantee that the minimum and maximum values of each attribute did not exceed the limits measured in the real sample. Each sample contained 2,000 pseudo-individuals. Using simulated data has the advantage that the modeller knows exactly how individuals behave; this allows detecting exactly the confounding effects that can occur. Having based the simulated experiment on real data allows us to generate behaviours that are feasible in reality and not completely abstract.

4. ESTIMATION RESULTS

Each dataset described in the previous section simulates a set of individuals with a well defined behaviour, known by the modeller and therefore perfectly reproducible. Using samples of simulated individuals in estimation allows to clearly detecting which effects can be confounded, as the modeller knows the underlying true behaviour. In particular, using each of the four samples described above, the following models were estimated:

ML1 a ML-RP model allowing for random taste heterogeneity specific to travel time by car
ML2 a ML-RP model allowing for random taste heterogeneity specific to travel time by public transport (bus and train)
ML3 a ML-EC model allowing for random heterogeneity specific to the car ASC
ML4 a ML-EC model allowing for correlation between bus and train
ML5 a ML-RP model allowing for generic random taste heterogeneity for travel time
MNL a MNL model with a step-wise linear function for travel time.

Table 1 show the results using datasets N.1 (models D1) and N.2 (models D2), while Table 2 the results using datasets N.3 (models D3) and N.4 (models D4). Firstly, it is interesting but worrying to note that all parameters (in all models) are highly significant whatever the specification used, be it the "right" specification (i.e. the same one used to generate the data) or any other "wrong" specification. Moreover, note that also the overall test of fit (the maximum log-likelihood) does not always help. In fact, in some cases there is not much difference between "right" and "wrong" specifications (compare for example models D1-ML1 and D1-ML3), but a "wrong" specification can even show a better fit (compare models D2-ML3 and D2-ML4). These results are worrying as in reality these tests are important for drawing conclusions about model results.

In particular, results from Table 1 clearly show the confounding effects among response heterogeneity, preference heterogeneity and correlation among alternatives and reveal the strong role of all these effects in accounting for omitted structures that differentiate alternatives. In fact, Dataset N.1 was generated under the assumption of randomness in travel time by car, but any random term that allows differentiating car and public transport alternatives seems to work well.

It is interesting to note that all models of type D1, except the "right" one (i.e. model D1-ML1) reproduce incorrectly a negative mean value for the car ASC; this confirms the importance of the ASC in judging model results (Cherchi and Ortúzar, 2005). If a wrong sign in the mean value of the ASC is an alarm of some sort in the utility specification, the opposite is not obviously true. In fact, all models D2 estimate the ASC correctly (right sign and value) even if a wrong utility function is specified. From models D2, it is also interesting to note that correlation between bus and train is not explained by the random travel time specific to the public transport (PT) alternatives (D2-ML2), while it is perfectly explained by the random travel time specific for the car alternative (D2-ML1), and by the random car ASC (D2-ML4). This result that might appear surprisingly at first glance is actually explained by the fact that differences in the travel time values for bus and train reduce the effect of the correlation due to specific PT randomness; conversely any randomness in the car (being this the only uncorrelated alternative) helps revealing the difference between the correlated alternatives (when correlation is wrongly not accounted for) and the alternative outside the nest. This result is not obvious and an interesting conclusion may be drawn since it seems evident that an underlying effect of correlation (and also of a RP specific for some alternative) could be to consider omitted structures that explain the diversity between car and public transport. Interestingly, thinking of correlation as a component of the error term or, even better, as unobserved heterogeneity, helps to understand the "real" role of correlation as a "proxy' for unobserved omitted structures.

The role of randomness in accounting for omitted structures is made clearer from the results in Table 2. Here modal choices in the dataset are generated according to a deterministic function for the travel time parameter. In both datasets (N.3 and N.4), therefore, the taste parameter for travel time is generic among alternatives and fixed over the sample. In particular, in Dataset N.3 the travel time parameter assumes two different values depending on whether the trip is longer or shorter than 10 minutes; and, interestingly, if this deterministic effect is omitted, the random heterogeneity in travel time (generic) appears highly significant (model D3-ML5). Instead, and as expected, neither the correlation effect (D3-ML4) nor the random parameter for the specific travel time by car (D3-ML1) are significant, because the omitted structure does not account for any similarities between groups of alternatives in this case.

Models D4 even show a more simple and common confounding effect due to omitted structures. In fact, models D4-ML1 and D4-ML2 were estimated omitting the frequency variable (which is defined only for the PT alternatives) and substituting this effect with specific random parameters for travel time by car (model D4-ML1) and for PT (model D4-ML2) respectively. The results discussed above are confirmed, as again the travel time by car parameter shows a significant randomness, while the PT travel time parameter does not. In fact randomness in PT travel time in our case is reduced by the strong differences in the travel time values of bus and train. Note that this effect suffices to prevent the appearance of confounding effects. If the heterogeneity in the PT travel time parameter was actually present in the data, this should have been detected in spite of the values of bus and train travel time.

This last result leads to an interesting consideration that might be of help when judging models. As confounding effects are implicit in any compensatory structure they can also appear among attributes (i.e. an attribute can act as proxy for an omitted structure); however, the less the variable used as proxy is constrained to the phenomenon the easier it is for it to account for any type of omitted structure; i.e. an EC structure is much likely to account for an omitted structure than a RP structure, because the former is not constraint to any attributes whose values are known by the modeller.

The possibility that all these confounding effects can appear (and actually do appear) in model estimation is certainly something to worry about, as it is very difficult to distinguish the real effect from the confounding one. However, the major problem is that confounding effects have a major impact in prediction, producing non-marginally different results.

Table 3 compares the market shares simulated and predicted with models D1, by showing the percent change in the aggregate share of mode *j* over the initial situation (do-nothing):

$$\Delta P_j = \frac{P_j - P_j^0}{P_j^0} \qquad (1)$$

where P_j^0, P_j are the aggregate probabilities of choosing mode *j* before (do-nothing) and after introducing the measure, calculated by sample enumeration.

As can be seen, models perform quite differently and sometimes errors are very large. It is interesting to note that, notwithstanding the wrong sign, model D1-ML3 performs quite well in predictions and sometimes (e.g. for a 30% reduction in travel time by car) even slightly better than model D1-ML1 (estimated with the same assumptions used to generate the data).

Table 3: Percentage Variation of the Predicted Demand

% variation in some attributes		True behaviour			D1-ML1			D1-ML2			D1-ML3		
		car	bus	train	car	bus	train	car	bus	train	car	bus	train
Train	-10%	-0,8%	-1,8%	6,9%	-1,2%	-1,5%	7,2%	-1,7%	0,9%	1,8%	-1,1%	-1,6%	7,0%
	-30%	-3,3%	-5,5%	21,9%	-3,2%	-5,1%	22,8%	-6,1%	2,1%	10,8%	-4,0%	-5,2%	22,8%
Car	-10%	10,6%	-2,3%	-2,4%	6,0%	-2,1%	-2,4%	3,1%	-1,7%	-0,9%	8,9%	-2,3%	-2,7%
	-30%	40,1%	-7,9%	-11,7%	22,9%	-7,2%	-9,0%	10,0%	-3,6%	-14,4%	30,4%	-7,6%	-9,5%

5. CONCLUSIONS

The ML model is characterized by a relatively simple structure and is supported by a well known and easy to follow theory. Moreover, its ability to account for any type of substitution patterns will make it probably the preferred model of the decade. A growing number of applications, both in research and applied work, have been reported and significant progress in its estimation has been made. Nevertheless, there are still many aspects of the ML formulation and interpretation of results that need to be explored.

In this paper we have examined the two typical structures of the ML model, the random parameters and the error components models, and highlighted some aspects that we think could have an impact on the interpretation of results. These refer particularly to the confounding effects among correlation, preference and response unobserved heterogeneity that these two structures generate due to the basic assumptions of any discrete choice model: (i) that individuals evaluate alternatives using compensatory rules; (ii) that the unobserved attributes might be accounted for by an additive random structure, and (iii) that only the difference between alternatives matters.

Using simulated data we performed several empirical tests and confirmed that in the ML model there is a clear trade-off between correlation and random heterogeneity. In general, our results clearly show that any unobserved randomness accounts for omitted structures, and this is not surprising as confounding effects are implicit in any compensatory structure and involve both the known and the unknown components. In particular, with the EC model we can account for any type of omitted structure as the EC structure does not have attributes associated to it. However, more worrying confounding effects might occur in the RP model especially in case of alternative specific random taste heterogeneity. For example, we found that a significant specific RP parameter for the travel time by car did not actually reveal variation in tastes for the travel time attributes, but correlation between the remaining PT alternatives. The same effect occurs also with a specific EC structure for the car alternative and reveals that, although formally different, correlation between PT alternatives and an EC structure specific for the remaining alternatives are actually complementary.

Our finding that the random parameter might not represent taste heterogeneity in that specific variable but other effects is certainly the most important result, especially if models need to be used in forecasting. This is particularly the case if the policies tested imply changes in the variable associated to the random parameter. In fact, although from a modelling point of view it is perfectly acceptable that there is a compensation between attributes and random terms, and good results are obtained with many different structures, we found that major problems arise in prediction as confounding effects produce non marginally different results.

Before using any model in forecasting, random heterogeneity must be analysed carefully and all the possible structures that might be confounded with it must be tested: correlation among alternatives, taste segmentation in the population and omitted structures not only in the alternative including the RP structure but also in the competing alternatives. We have found no papers carrying on these further analyses after founding significant heterogeneity in tastes. We also suggest, in each specification, to look carefully at the estimates of the alternative specific constants. A wrong sign in the mean value of an ASC constitutes an alarm of some sort in the utility specification, the opposite is not obviously true.

ACKNOWLEDGEMENTS

The simulation programme used in this research is based on work done during a stay of the first author at the Centre for Transport Studies, University College London; we want to thank Prof. John Polak for his help and guide during this work. We are grateful to FONDECYT for having contributed to this line of research through Projects 1020981 and 1050672.

REFERENCES

Bath, C. and S. Castelar. (2002) A unified mixed logit framework for modelling revealed and stated preferences: formulation and application to congestion pricing analysis in the San Francisco Bay Area. **Transportation Research 36B**, 593-616.

Ben-Akiva, M.E. and S.R. Lerman. (1985) **Discrete Choice Analysis: Theory and Application to Travel Demand**. The MIT Press, Cambridge, Mass.

Cantillo, V. and J. de D. Ortúzar. (2005) A semi-compensatory discrete choice model with explicit attribute thresholds of perception. **Transportation Research 39B**, 641-657.

Cherchi, E. and J. de D. Ortúzar. (2002) Mixed RP/SP models incorporating interaction effects: modelling new suburban train services in Cagliari. **Transportation 29**, 371-395.

Cherchi, E. and J. de D. Ortúzar. (2005) On fitting mode specific constants in the presence of new options in RP/SP models. **Transportation Research A** (in press).

Galilea, P. and J. de D. Ortúzar. (2005) Valuing noise level reductions in a residential location context. **Transportation Research 10D**, 305-322.

Hensher, D.A. (1998) Extending valuation to controlled value functions and non-uniform scaling with generalised unobserved variances. In T. Gärling, T. Laitila and K. Westin (eds.), **Theoretical Foundations of Travel Choice Modelling**. Elsevier, Amsterdam.

Hensher, D.A. and W.H. Greene. (2003) The mixed logit model: the state of practice. **Transportation 30**, 133-176.

McFadden, D. and K. Train. (2000) Mixed MNL models for discrete response. **Journal of Applied Econometrics 15**, 447-470.

Munizaga, M. and R. Alvarez. (2001) Mixed logit vs. nested logit and probit models. *5th* **Tri-Annual Invitational Choice Symposium. Workshop: Hybrid Choice Models, Formulation and Practical Issues**. Asilomar, Texas.

Swait, J.D. and A. Bergantino. (2000) Distinguishing taste variation from error structure in discrete choice data **Transportation Research 34B**, 1-15.

Train, K.E. (2003) **Discrete Choice Methods with Simulation**- Cambridge University Press, Cambridge.

Williams, H.C.W.L. and J. de D. Ortúzar. (1982) Behavioural theories of dispersion and the mis-specification of travel demand models, **Transportation Research 16B**, 167-219.

Table 1
Model estimation results with simulated data: random parameter and correlation

	set N.1 (models D1)					taset N.2 (models D2)			
Attributes	**D1-ML1**	**D1-ML2**	**D1-ML3**	**D1-ML4**	**D1-MNL**	**D2-ML1**	**D2-ML2**	**D2-ML3**	**D2-ML4**
Travel time PT (mean)	-0,1608 (-14,86)	-0,1037 (-10,67)	-0,1644 (-14,92)	-0,1645 (-14,89)		-0.0951 (-12.00)	-0.0931 (-9.08)	-0.1231 (-9.36)	-0.1223 (-9.31)
Travel Time PT (st,dev,)		0,0388 (2,70)					0.0169 -1.18		
Travel time Car (mean)	-0,7394 (-7,84)	-0,1690 (-9,97)	-0,3900 (-8,48)	-0,3928 (-8,40)		-0.1061 (-10.31)	-0.1039 (-10.28)	-0.446 (-7.62)	-0.1437 (-9.31)
Travel Time Car (st,dev,)	0,4403 (7,36)					0.0443 -2.22			
Travel time (first 10 min)					-0,0481 (-2,60)				
Travel Time (minutes after the 10th)					-0,0956 (-17,40)				
Walking time	-0,3287 (-15,84)	-0,2179 (-14,01)	-0,3352 (-15,70)	-0,3353 (-15,73)	-0,1995 (-20,95)	-0.2476 (-13.34)	-0.2356 (-14.54)	-0.3114 (-9.92)	-0.3099 (-9.87)
Cost	-0,7404 (-5,46)	-0,3902 (-4,47)	-0,7954 (-5,49)	-0,7907 (-5,45)	-0,3437 (-4,67)	-0.4802 (-5.78)	-0.4485 (-6.07)	-0.6279 (-5.49)	-0.6213 (-5.48)
Frequency	0,9001 (13,92)	0,5875 (10,28)	0,9350 (14,24)	0,9344 (14,24)	0,5139 (16,56)	0.6544 -14.23	0.6281 -14.74	0.8368 -10.22	0.8317 -10.16
ASC_car (mean)	14757 (2,66)	-0,7542 (2,56)	-13597 (-2,43)	-13417 (-2,41)	-17208 (-8,46)	2.1996 -6.5	2.1426 -6.64	3.0143 -5.96	2.9903 -5.94
ASC_car (st, dev,)			-52756 (-8,61)					1.8366 -4.91	
K_train	-0,8177 (-4,97)	-0,5736 (-2,56)	-0,8023 (-4,75)	-0,8043 (-4,75)	-0,7769 (-63,85)	-0.6676 (-3.97)	-0.6415 (-3.93)	-0.7024 (-3.53)	-0.7034 (-3.55)
Error comp (equal for PT)			--	-52277 (-8,65)				--	1.8137 -4.8
L(max)	-1069.9	-1157.6	-1095.5	-1095.6	-1170.9	-1023.9	-1024.9	-1018.9	-1019.2
Number of observations	2000	2000	2000	2000	2000	2000	2000	2000	2000

Table 2
Model estimation results with simulated data: deterministic utilities

	Dataset N.3 (models D3)				Dataset N.4 (models D4)		
Attributes	**D3-MNL**	**D3-ML1**	**D3-ML4**	**D3-ML5**	**D4-MNL**	**D4-ML1**	**D4-ML2**
Travel Time (mean)			-0.1804 *(-17.44)*	-0.2067 *(-14.96)*	-0.1496 *(-19.91)*		
Travel Time (st.dev.)				0.0660 *(4.64)*			
Travel time PT (mean)		-0.1705 *(-14.51)*				-0.1110 *(-14.00)*	-0.1157 *(-8.81)*
Travel Time PT (st.dev.)							0.0248 *(1.70)*
Travel time CAR (mean)		-0.2140 *(-12.58)*				-0.1203 *(-12.24)*	-0.1228 *(-10.35)*
Travel Time CAR (st.dev.)		-0.0235 *(-0.72)*				-0.0365 *(-2.06)*	
Travel time (first 10 min)	-0.8197 *(-15.40)*						
Travel Time (min. after the 10^{th})	-0.1519 *(-16.99)*						
Walking time	-0.3310 *(-18.98)*	-0.2924 *(-14.55)*	-0.2868 *(-17.10)*	-0.3101 *(-16.58)*	-0.32 *(-20.19)*	-0.2565 *(-15.74)*	-0.2553 *(-14.96)*
Cost	-0.8238 *(-9.55)*	-0.7214 *(-8.06)*	-0.6910 *(-8.64)*	-0.7409 *(-8.80)*	-0.7884 *(-9.84)*	-0.6217 *(-7.58)*	-0.6159 *(-7.60)*
Frequency	0.9180 *(18.85)*	0.8410 *(14.22)*	0.8291 *(15.60)*	0.8908 *(16.34)*	0.8519 *(19.22)*		
ASC_car mean	3.0128 *(10.61)*	3.5802 *(8.93)*	2.7337 *(9.39)*	2.9516 *(9.85)*	2.6335 *(10.02)*	-1.8403 *(-6.94)*	-1.8432 *(-6.66)*
ASC_car St. Dev.			0.1474 *(0.19)*				
K_train	-0.6860 *(-3.67)*	-0.3077 *(-1.85)*	-0.4039 *(-2.53)*	-0.4543 *(-2.60)*	-0.7354 *(-4.19)*	-2.5391 *(-14.54)*	-2.5001 *(-15.74)*
L(max)	-699.4	-760.7	-767.0	-762.7	-753.5	-1015.87	-1016.04
Number of observations	2000	2000	2000	2000	2000	2000	2000